Number:

Problem	Points	Score
1 (a)	10	
1 (b)	10	
2 (a)	10	
2 (b)	10	
3 (a)	10	
3 (b)	10	
4 (a)	10	
4 (b)	10	
5	10	
6	10	
Total	100	

Notes:

1. The exam is closed books and notes. You are allowed three 8 1/2” x 11” double-sided sheet of notes.

2. Please indicate clearly your answer to the problem by some form of highlighting (underlining).

3. Your solutions must be legible and easy to follow. If I can’t read it or understand it, it is wrong. Random scribbling will not receive credit.

4. Please show ALL work. Answers with no supporting explanations or work will be given no credit.

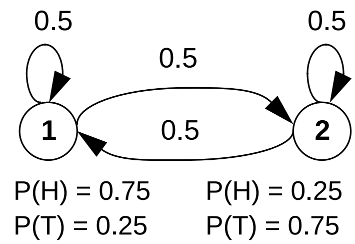
5. Several problems on this exam are fairly open-ended. Since the evaluation of your answers is obviously a subjective process, we will use a market place strategy in determining the grade. Papers will be rank-ordered in terms of the quality of the solutions, and grades distributed accordingly.

1. Consider the signal $x(t) = A\sin(2\pi f_o t) + (A/2)\sin(2\pi(2f_0)t)$, where $f_0 = 500\text{ Hz}$.

(a) Assume this signal is sampled at $f_s = 1\text{ kHz}$. Compute the linear prediction coefficients for a fourth order model using the minimum number of samples necessary to get an exact estimate of the coefficients. Explain carefully how you determined this number of samples.

(b) Suppose you repeated this calculation for $f_s = 2\text{ kHz}$. Do the values of the coefficients change? Why? Explain.

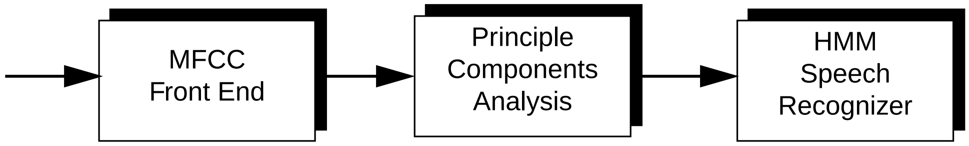
2. A system ("black box") outputs the sequence "HT". Assume the probability of starting in each state (initial probabilities) are equal.



(a) Use the Baum Welch algorithm to determine the posterior probability of this observation sequence (probability of the data given the model).

(b) Use the Viterbi algorithm to compute the optimal state sequence and posterior probability. Explain why the answers to (a) and (b) differ. Support your argument by comparing this simple problem to what is done in a typical speech recognition system.

3. Consider the following speech recognition system:



(a) Describe the steps performed in computing a standard MFCC feature analysis of the signal. Include delta and delta-delta coefficients. Justify each step based on knowledge about the speech signal. Be as specific and detailed as possible.

(b) Explain the merits (or drawbacks) of doing a global PCA transformation prior to applying the features to the HMM recognizer.

4. Recall our standard block diagram model of a speech recognition system (front end, acoustic modeling, language modeling, and search).

(a) Describe what aspects of the speech recognition statistical model are implemented by the search module.

(b) Consider two applications: credit-card number recognition on clean data, and conversational speech transcription of telephone data. We discussed two major paradigms for search: breadth-first and depth-first. Which of these search algorithms is most appropriate for each application. Justify your answer in terms of accuracy and computational complexity.

5. You flip a coin and observe the following data: “HHHHHHTTTTTTHHHHTTTT”. After repeated trials with the same coin, you observe the same trends — sequences of four or more heads or tails are highly probable, and you never see a single H or T (H or T are observed in runs of two or more). Design a speech recognition system based on principles of maximum likelihood estimation that best models this data.

6. Zheng has run an experiment on noisy cellular phone data collected in automotive environments using a hands-free microphone. This table to the right summarizes performance in terms of word error rate. The baseline condition is a standard LVCSR system (MFCCs, time synchronous Viterbi search, etc.). Exp. no. 2 uses a noise cancellation technology that is known to significantly increase the perceptual quality of the audio data as well as the signal to noise ratio. In exp. no. 2, noise cancellation is performed only on the evaluation data, and not on the training data. In exp. no. 3, noise cancellation is performed on both training and evaluation data. Assume there are no bugs in these experiments.

System	Word-Internal	Cross-Word
Baseline	46.7	44.3
Noise Cancellation w/o No Retraining	49.3	46.9
Noise Cancellation w/ Retraining	50.9	50.9

Zheng argues that these results make sense and that he would have predicted this outcome *before* ever running these experiments. I am not so sure. Provide an analysis that explains why retraining worsened performance. This seems impossible given what we learned about EM, MLE, and other such techniques used in this system.